Published in final edited form as:

Alzheimers Dement. 2016 January ; 12(1): 2–10. doi:10.1016/j.jalz.2015.05.020.

Genome-wide linkage analyses of non-Hispanic White families identifies novel loci for familial late-onset Alzheimer's disease

Brian W. Kunkle[1], **James Jaworski**[1], **Sandra Barral**[2,3,4], **Badri Vardarajan**[2,3,6], **Gary W. Beecham**[1], **Eden R. Martin**[1], **Laura S. Cantwell**[7], **Amanda Partch**[7], **Thomas D. Bird**[8,9], **Wendy H. Raskind**[9,10], **Anita L. DeStefano**[11], **Regina M. Carney**[1,12], **Michael Cuccaro**[1,13], **Jeffrey M. Vance**[1,13], **Lindsay A. Farrer**[11,14], **Alison M. Goate**[15], **Tatiana Foroud**[16], **Richard P. Mayeux**[2,3,4,17], **Gerard D. Schellenberg**[7], **Jonathan L. Haines**[18], and **Margaret A. Pericak-Vance**[1,13]

[1]John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA

[2]The Taub Institute of Research on Alzheimer's Disease, College of Physicians and Surgeons, Columbia University, New York, NY, USA

Corresponding Author: Margaret Pericak-Vance, PhD, John P Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, 1501 NW 10th Ave, Miami, FL 33136, Phone: 877-686-6444, mpericak@med.miami.edu.

AUTHORS CONTRIBUTIONS:

Study Concept and Design: Pericak-Vance, Kunkle, Haines, Mayeux

Acquisition of Data: Pericak-Vance, Mayeux, Schellenberg, Bird, Raskind, Haines, Carney, Cuccaro, Vance

Analysis or Interpretation of Data: Kunkle, Jaworski, Barral, Vardarajan, Pericak-Vance, Mayeux, Haines, Beecham, Martin, Farrer, Schellenberg

Drafting of Manuscript: Kunkle, Pericak-Vance, Barral, Haines

Critical Revision of Manuscript for Important Intellectual Content: Pericak-Vance, Haines, Kunkle, Farrer, Mayeux, Schellenberg, Barral, Goate, DeStefano, Foroud

Obtaining Funding for Study: Pericak-Vance, Mayeux, Schellenberg, Haines

Administrative, Technical or Material Support: Cantwell, Partch

Study Supervision: Pericak-Vance, Haines, Mayeux

DATA SHARING

All data from the analyses in this manuscript, including quality control documentation, GWAS array data and phenotype data for each family, and linkage analyses results, is available for download at The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site's ADSP website (https://www.niagads.org/adsp/content/home). Applicants must submit a data access request to dbGaP. Applications are reviewed by the ADSP Data Access Committee (DAC) and the NIAGADS Data Use Committee (DUC).

ADSP Data Sharing Agreement

ADSP phenotype and sequence data are made available to the research community at large in keeping with the NIH Genomics Data Sharing Policy http://gds.nih.gov/. NIA has established the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) as a national genetics data repository in order to facilitate access by qualified investigators to genotypic and phenotypic data for the study of the genetics of late-onset Alzheimer's disease. NIAGADS is working in partnership with dbGaP (ADSP at dbGaP) to provide ADSP data to the research community. Data can be requested either from dbGaP or NIAGADS. Instructions for application for ADSP data and an explanation of the review process can be found at: ADSP at dbGaP and NIAGADS ADSP Application Instructions.

The ADSP has in place a memorandum of understanding: https://www.niagads.org/sites/all/public_files/ADSPdocs/ADSP-MOU.pdf. In the spirit of the clear benefit that ensues from converting such data sets into community resources as rapidly as possible, and in keeping with community expectations for the use of unpublished genome sequence data, it is expected for the first phase of the study called the Discovery Phase, that users of the data will withhold publication until the producers of the data have published their findings. ADSP participants will publish their data in an expeditious fashion in at least one major paper reporting the results of the ADSP to be jointly submitted by all of the members.

[3]The Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY, USA

[4]The Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA

[5]The Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY, USA

[6]Department of Psychiatry and Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA

[7]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[8]Department of Neurology, University of Washington, Seattle, WA, USA

[9]Department of Medicine, University of Washington, Seattle, WA, USA

[10]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

[11]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

[12]Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA

[13]Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA

[14]Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, and Epidemiology, Boston University Schools of Medicine and Public Health, MA, USA

[15]Department of Psychiatry and Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA

[16]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

[17]The Department of Epidemiology, School of Public Health, Columbia University, New York, NY, USA

[18]Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH, USA

Abstract

INTRODUCTION—Few high penetrance variants that explain risk in Late-onset Alzheimer's disease (LOAD) families have been found.

METHODS—We performed genomewide linkage and identity-by-descent (IBD) analyses on 41 non-Hispanic Caucasian families exhibiting likely dominant inheritance of LOAD, and having no mutations at known familial AD loci and a low burden of *APOE* ε4 alleles.

RESULTS—Two-point parametric linkage analysis identified 14 significantly linked regions, including three novel linkage regions for LOAD (5q32, 11q12.2-11q14.1 and 14q13.3), one of

which replicates a genomewide association LOAD locus, the *MS4A6A-MS4A4E* gene cluster at 11q12.2. Five of the 14 regions (3q25.31, 4q34.1, 8q22.3, 11q12.2-14.1 and 19q13.41) are supported by strong multipoint results (LOD*≥1.5). Non-parametric multipoint analyses produced an additional significant locus at 14q32.2 (LOD*=4.18). The 1-LOD confidence interval for this region contains one gene, *C14orf177*, and the miRNA *Mir_320*, while IBD analyses implicates an additional gene *BCL11B*, a regulator of brain-derived neurotrophic signaling, a pathway associated with pathogenesis of several neurodegenerative diseases.

DISCUSSION—Examination of these regions following whole genome sequencing may identify highly penetrant variants for familial LOAD.

Keywords

Non-Hispanic White; Late Onset Alzheimer's Disease; linkage; high penetrance; identity-by-descent; familial; genetics

1. BACKGROUND

While more than two dozen loci that contribute to late-onset Alzheimer disease (LOAD) have been identified [1], few genes with highly penetrant rare variants (e.g. *APP*, *PSEN1* and *PSEN2* in early-onset familial AD [2]) that explain risk in families heavily burdened with LOAD have been found. It is likely that rare variants contribute to complex disease, however [3], and recent reports implicating rare variants in *PLD3*, *APP* and *TREM2* [4–7] support their involvement in both sporadic and familial LOAD. Identification of additional rare mutations driving genetic risk in familial LOAD will help in defining new pathways for therapeutic and preventive treatments.

Linkage analyses in large multiplex pedigrees is a robust approach for identifying disease loci in the presence of allelic heterogeneity, and thus can be valuable for targeting regions for sequencing studies [8]. To identify genomic regions likely to contain rare (MAF ≤ 0.01) and low-frequency (0.01 ≥ MAF ≤ 0.05) LOAD risk and possibly protective genetic variants, a large number of well-characterized families were screened for inclusion in a linkage scan. The selected extended families are uniquely suited for discovery of genomic regions containing high penetrant Alzheimer's disease variants. We performed extensive parametric two-point and non-parametric multipoint linkage analysis on 385 individuals in 41 non-Hispanic Caucasian (NHC) families. Loci identified through this study can help prioritize regions of the genome for analyses of whole exome or whole genome sequence data from NHC LOAD families or case-control cohorts.

2. METHODS

2.1. Study samples

The 42 NHC families selected for linkage analyses are from five collections assembled by investigators at The University of Pennsylvania (8 Families), The University of Miami (12 Families), Case Western University (1 Family), the National Institute on Aging Late-Onset Alzheimer's Disease (NIALOAD) family study (17 Families), and the National Cell Repository for Alzheimer's Disease (NCRAD) (4 Families). Detailed descriptions of the

ascertainment and evaluation of subjects in these cohorts have been provided elsewhere [9–11]. To maximize the probability of detecting segregating novel rare variants, we developed several selection criteria including: (1) having four or more affected individuals with genomic DNA samples (2) exhibiting likely dominant inheritance of LOAD; (3) free of known mutations at established AD/FTD Mendelian loci (*APP*, *PSEN1*, *PSEN2*, *MAPT*, or *GRN*); and (4) reduced representation of the *APOE* ε4 allele. Criterion number 4 included prioritized selection of families with APOE ε2/ε2, ε2/ε3 and ε3/ε3 affected individuals (requiring at least one affected family member without any APOE ε4 allele and any affecteds with a single APOE ε4 must have age-at-onset (AAO) < 72). 385 individuals in the 41 NHC families ultimately analyzed (3-11 cases per pedigree) had genotyping data available and were included in the present analyses. 75.6% of families (31 of 41) have at least one autopsy confirmed LOAD case (Table 1).

2.2. Genotyping and quality control procedures

Genome-wide single-nucleotide polymorphism (SNP) genotyping was performed on several different platforms across the study cohorts, including the Illumina HumanHap 550, Illumina 1M, HumanOmniExpress, HumanOmniExpress Exome, and HumanOmni2.5 arrays. A call rate threshold of 98% was applied and the data were then merged to form a final linkage dataset for analysis. SNPs were only included in the analysis if they were present in at least 60% of samples. 319,409 SNPs were selected for analysis and aligned to the Rutgers Map v.3 [12]. Among this group of SNPs, 26,959 were excluded because the minor allele frequency (MAF) was less than 0.05 and/or the genotype distribution differed significantly ($P < 10^{-6}$ in controls) from Hardy-Weinberg equilibrium. An additional 919 SNPs not present in the HapMap CEU dataset were removed, reducing the number of SNPs available for analysis to 291,531 SNPs. More than three-fourths of these SNPs (77%; n = 225,250 SNPs) were present in 90 or more percent of samples. Checks for relatedness, Mendelian inconsistencies and gender based on X chromosome heterozygosity were performed using PLINK [13]. One sample was dropped due to Mendelian inconsistencies and one duplicate sample was removed. Principal components analysis using Eigenstrat [14] identified a family clustering with African American HapMap samples (eFigure 1). NIALOAD confirmed the family's African American ancestry through recontact with the ascertainment site. This family was removed from the analyses resulting in 41 families included in the linkage analyses reported here.

2.3. Statistical analyses

Autosomal and X-chromosome linkage analyses were performed using Merlin [15] and included parametric two-point affecteds-only and age-dependent penetrance models, and a non-parametric multipoint analysis. Parametric multipoint analysis was performed on significant overlapping regions between the families in this report and a companion analysis in Hispanics (Barral et al. in this issue). The package MINX (Merlin in X) was used for analysis of X-chromosome SNPs. Heterogeneity LOD (HLOD) models were applied to the two-point analyses to allow for detection of linkage in the presence of locus heterogeneity [16]. Whittemore and Halpern NPL-pair and NPL-all statistics [17], and Kong and Cox linear model logarithm of odds (LOD*) scores [18], were calculated for the non-parametric multipoint analysis.

Power analyses using SIMLINK[19] on the 41 families in the linkage analysis, with a dominant model and disease allele frequency of 0.001, showed we have >80% power to generate a LOD > 3 for a fully informative (alpha = 1) age-penetrance model with marker locus allele frequencies equal to 0.2 (MAXLOD = 5.62) and 0.4 (MAXLOD = 6.86). Using these same paramaters, the affecteds-only model has >80% power to generate a LOD > 2 with a marker locus frequency of 0.4 (MAXLOD = 3.61) and 0.71% power to produce a LOD > 2 with a marker locus frequency of 0.2 (MAXLOD = 3.00). Using a heterogeneity model (alpha = 0.5) reduced power to generate a LOD > 2 to 41% and 18% for the age-penentrance and affecteds-only models respectively (eTable 1).

Parameters for the parametric two-point models assumed dominant inheritance, a disease allele frequency of 0.001 and penetrances of 0.01, 0.90, and 0.90 (representing NN, NA, AA genotypes respectively). Age-dependent penetrances used in the analysis are listed in eTable 2. Two-point parametric analysis utilized all SNPs for each of the analyses. The non-parametric multipoint scan included a linkage disequilibrium (LD)-pruned set of 119,555, SNPs common to all genotype platforms. LD pruning was done using the independent pairwise LD pruning option in Plink (default settings). Mean distance between markers for the set of non-parametric multipoint markers is 4.55 cM. As some pedigrees were too large for MERLIN to perform nonparametric linkage analysis, uninformative family members (based on an individual's position in the pedigree and/or absence of genotyping) were trimmed before performing analyses using the program PowerTrim [20]. Allele frequencies for all SNPs were based on CEU HapMap data [21].

A significance threshold of HLOD ≥ 3.5 was set for the parametric two-point linkage scans taking into account testing of two separate parametric models. This is above the Lander and Kruglyak recommendations for significance (LOD ≥ 3.3; P value = 4.9×10^{-5}) in LOD score analyses of dense marker genome-wide linkage scans [22], and approximates a level suggested by Camp and Farnham for testing of two independent two-point models [23]. Multi-point significant and suggestive linkage thresholds were defined by LOD ≥ 3.60, $P = 2.2\times10^{-5}$ and LOD ≥ 2.20, $P = 7.4\times10^{-4}$, respectively [22]. Linkage regions were considered independent if the locations of their peak HLOD or LOD* scores were separated by >20 cM. Linkage peaks were considered concordant with previous linkage peaks or linkage peaks reported in the companion Hispanic linkage analysis (Barral et al. 2014 in this issue), if they were ≤10 cM apart.

Follow-up analyses of significant multipoint results included haplotype segregation analysis and examination of overlapping identity-by-descent (IBD) sharing segments for families with a maximized within family LOD ≥ 0.59, corresponding to a nominal P value of 0.05, and 100% IBD sharing among all affecteds in the family. IBD segments were determined by estimating haplotypes in MERLIN followed by identification of IBD sharing regions using Olorin [24].

3. RESULTS

3.1. Dataset Characteristics

The selected families have an average of 8 affected individuals (range: 4-14), with an average of 5 genotyped affected individuals per family (range: 3-11) (Table 1). Mean AAO in affecteds was 75 ± 9.04, compared to a mean age of unaffecteds at last evaluation of 66 ± 12.29. *APOE* ε2, ε3, and ε4 frequency in affecteds (genotyped and ungenotyped) was 3%, 70%, and 27% respectively. This compares to an ε4 frequency of 38% in Caucasian LOAD individuals and 14% for controls [25], confirming the selection of LOAD families with reduced representation of the *APOE* ε4 allele. *APOE* genotype frequencies for affecteds were: ε2/ε2 = 0.5%, ε2/ε3 = 4.1%, ε2/ε4 = 0.9%, ε3/ε3 = 43.2%, ε3/ ε4 = 48.6%, and ε4/ ε4 = 2.7%.

3.2 Linkage Results

Two-point parametric linkage analysis identified 24 SNPs across 14 autosomal regions surpassing a significance level of HLOD ≥ 3.5 (P value = 3×10^{-5}) (Table 2, Figures 1 and 2). Nine of the 14 regions generated HLOD scores that exceeded 4.0, including 3q25.31 (HLOD = 4.15), 3q27.3 (HLOD = 4.58), 4p34.3 (HLOD = 4.46), 5q32 (HLOD = 4.10), 7p21.2 (HLOD = 4.19), 9p22.1 (HLOD = 4.21), 11q13.4 (HLOD = 4.74), 16q12.1 (HLOD = 4.05) and 19q13.41 (HLOD = 4.76). As expected based on the family selection criteria for exclusion of clustering of affected subjects who were predominantly *APOE* ε4, we did not observe linkage to the *APOE* locus. Age-dependent penetrance HLOD scores were generally lower than our affecteds-only HLOD scores. A majority of SNPs generated HLOD scores with alpha values equal to 1.0, suggesting modest contributions to individual loci by each family. No significant results were observed for the overall analyses on the X-chromosome, however two families had nominally significant LOD scores of 1.39 and 1.14 at Xq28 (rs5963398), the location of the highest HLOD score on the X-chromosome (HLOD = 2.07).

Non-parametric multipoint analyses identified an additional significant region at 14q32.2 (LOD* = 4.18) (Figure 2). Two-point results for this region were also supportive of linkage (HLOD = 2.82 at rs9323997; affecteds-only model). One other region produced suggestive multi-point results, 4q34.1-4q34.3 (LOD = 2.40), and was supported by a significant two-point linkage (HLOD = 4.46; affecteds-only model). Please see eTable 3 for the 1-LOD region and genes within this region. No significant multipoint results were observed on the X-chromosome.

3.3. Localization of region at 14q32.2

The resulting 1-LOD region [16] at 14q32.2 is a 0.78 Mbp segment between map positions 98.81 Mbp and 99.59 Mbp, and contains one gene (*C14orf177*) and one microRNA (miRNA), *Mir_320* according to the UCSC Genes Track [26]. The pseudogene ribosomal protein L3 psudeogene 4 (*RPL3P4*) also locates to this segment according to the Gencode database (version 19) [27]. Linked pedigree IBD sharing analyses among the four nominally significantly linked pedigrees (LOD* ≥ 0.59) isolated a 0.40 Mbp segment containing one gene, *BCL11B*, located just outside the 1-LOD region, and *Mir_320* (Figure 2). Haplotype segregation of these 4 pedigrees illustrated using the software program Progeny (Progeny

Software LLC, Delray Beach, FL www.progenygenetics.com) is shown in eFigures 2-5. Merlin software was used for haplotype construction and inference.

4. DISCUSSION

We report 14 significant two-point linkage regions (HLOD ≥ 3.5) and one significant multipoint region (LOD* ≥ 3.6) identified by analysis of 41 multiplex LOAD families that were selected on the basis of the absence of variants at known risk loci, apparent dominant inheritance of disease, and little evidence for association of LOAD with the *APOE* ε4 allele. Five of the 14 two-point regions are supported by multipoint results within a 1-LOD unit confidence interval with a minimum LOD* ≥ 1.5 (3q25.31, 4q34.1, 8q22.3, 11q12.2-14.1 and 19q13.41). The finding of numerous significantly linked loci instead of a few shared loci suggests that there is substantial locus heterogeneity within familial LOAD.

Three of the 14 two-point loci are novel linkage regions for LOAD (5q32, 11q12.2-11q14.1 and 14q13.3). The 11 other loci overlap previously reported LOAD regions including five significant loci (3q25.31, 7p21.2, 9p22.1, 11q12.2-11q14.1 and 19q13.41) also showing significant linkage (two-point HLOD ≥ 3.3, multipoint LOD* ≥ 3.6) or association ($p \geq 5 \times 10^{-8}$) in at least one previous report (Table 2). These include two loci reported in the largest LOAD GWAS to date, namely the 11q12.2-11q14.1 locus which overlaps the *MS4A6A-MS4A4E* association region and the 19q13.41 locus which contains *CD33* [1]. One additional GWAS locus (CR1 at 1q32.2) is near a significantly linked marker at 1q32.3. A check of our linked SNPs in these regions for association with LOAD in the IGAP GWAS summary statistics found no significant associations after correction for multiple testing ($P \leq 5 \times 10^{-8}$) (eTable 4) [1]. The 7p21.2 and 9p22.1 loci replicate significant linkage from reports that contain families used in the present analyses (Table 2). The number of pedigree members and the phenotypes and genotypes from the previous reports have been continuously expanded and updated since their previous linkage reports however, and these updates are most likely increasing power to localize linkage in this current report.

The locus at 14q32.2 is arguably our most robust result given its significant multi-point LOD* score supported by suggestive two-point scores. This locus is a considerable distance (~33 cM) from the *PSEN1* locus at 14q24. Suggestive linkage (Two-point LOD = 2.60) has been reported at 14q32 in an age-at-onset linkage analysis in Hispanic LOAD families [28]. The 1-LOD limit identified a region 0.78 Mbp in length containing *C14orf177* and *Mir_320*, both of which have some support for involvement in dementia-related disease and processes. *C14orf177* for instance, has been associated with risk for amyotrophic lateral sclerosis [29] and lipoprotein cholesterol levels [30], while members of the mir-320 microRNA family are significantly altered in sporadic AD brains [31] and associated with both neurite outgrowth [32] and neurodegeneration [33]. Evidence for genomic features with regulatory potential such as several ESTs and lincRNAs also exists in the region (based on UCSC genome browser data)[26], including one lincRNA in particular, TCONS_12_00008237, which is highly expressed in brain [34,35].

The region narrowed by IBD analysis in the subset of pedigrees most likely to be linked to 14q32.2 also includes *BCL11B*, which is a transcription factor and regulator of BDNF

signaling [36], a pathway associated with pathogenesis of several neurodegenerative diseases, including LOAD [37]. *BCL11B* is predominantly expressed in striatal neurons and may play an important role in adult neurogenesis [38], a process that when dysregulated may lead to AD [39]. *BCL11B* is thought to primarily reduce BDNF signaling [36], consistent with observations that BDNF serum and expression levels are decreased in AD [37,40,41], and high BDNF levels protect against AD [42,43]. A recent neuroimaging study implicated a role for BDNF in cognitive decline in LOAD patients [44], although AD genetic association studies of BDNF yielded conflicting results [40,45]. Finally, it has been suggested that BDNF-based drugs might be effective therapies for AD and other neurodegenerative diseases [46], with targeting of *BCL11B* interactions with BDNF even being suggested as a feasible therapeutic approach to elevate BDNF signaling in neurodegenerative drug development [36].

The accompanying study in Caribbean Hispanics by Barral et al [47] found strong evidence for linkage and association near one of our significant loci, 11q12.2-11q14.1. As noted above, the chromosome 11q12.2-11q14.1 locus is also a significant LOAD GWAS locus from Lambert et al. 2013 [1]. Parametric affecteds-only analysis of this region produced a multipoint peak LOD* of 1.18. An alpha of 0.17 suggests that only a small number of pedigrees are potentially segregating a variant in this region. A combined association and linkage analysis of all markers in this region using CAPL [48] produced no suggestive or significantly associated SNPs (data not shown). One possible explanation for this finding is lack of power for combined linkage and association analysis using these data. Several other significant loci from the Barral et al. study had suggestive linkage in our analyses, including 3q13.31 (HLOD = 3.31), 3q22.3 (HLOD = 3.18), 6q25.3 (HLOD = 3.02), 7p14.3 (HLOD = 3.05), and 14q12 (HLOD = 3.31) (eTable 5). The 3q22.3 region was previously reported as a potential locus for LOAD in a linkage study of a family containing four relatives with LOAD but without tau pathology (LOD = 4.1) [49], and in a genome-wide linkage study of Dutch families (LOD = 4.3; HLOD = 4.4) [50].

In summary, we report 15 significant regions for linkage, including novel evidence for linkage at 5q32, 11q12.2-11q14.1 and 14q13.3. Several of our regions overlap significant loci from previous LOAD analyses, including GWAS regions at *MS4A6A*-*MS4A4E* and *CD33*. Our strong multipoint result at 14q32.2 is particularly interesting, as it localizes to a region with a limited amount number of genomic candidates, most with plausible links to dementia-related processes and disease.

The 41 families included in these analyses are undergoing whole-genome sequencing (WGS) as part of the National Institute of Health's Alzheimer's Disease Sequencing Project (ADSP) [51]. ADSP WGS variants located in these linkage regions will be primary candidates for examination as contributing to risk or protection for LOAD. Analyses planned by the ADSP to identify these variants include: 1) combined linkage and association analyses, and 2) filtering for rare, damaging variants in shared familial segments. An ADSP replication phase will follow to confirm and validate candidate loci from the discovery phase.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS/CONFLICTS/FUNDING SOURCES

We thank Drs. Eric Boerwinkle (University of Texas Health Science Center), Josh Bis (University of Washington) and Benjamin Neale (Harvard Medical School) for their help in selection of the families for the analyses presented in this manuscript. We are grateful to the families and staff who participated in this study. This work was supported by National Institute of Health grants R01 AG027944 (MPV), R01 AG019085, R01 AG028786-02 (MPV), RC2AG036528 (MPV; JLH), R37AG015473 (RM), U01AG032934 (GS), R01-AG025259 (LF), P30-AG13846 (LF), 1R01 NS069719-01 (WHR), the Alzheimer Disease Research Center Genetics Core grant 5P50-AG008702-25 (RM), VA Medical Research Funds (TB), the Alzheimer's Association grant IIRG09133827 (MPV), Department of Veterans Affairs Merit Review Award (TDB; WHR), and the VA Puget Sound Health Care System Geriatric Research, Education and Clinical Center (TDB). None of the authors has a conflict of interest.

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i–Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC–10–196728.

REFERENCES

1. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013; 45:1452–8. doi:10.1038/ng.2802. [PubMed: 24162737]
2. Bird TD. Genetic aspects of Alzheimer disease. Genet Med. 2008; 10:231–9. doi:10.1097/GIM.0b013e31816b64dc. [PubMed: 18414205]
3. Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. Population genetics of rare variants and complex diseases. Hum Hered. 2012; 74:118–28. doi:000346826. [PubMed: 23594490]
4. Ramirez A, van der Flier WM, Herold C, Ramonet D, Heilmann S, Lewczuk P, et al. SUCLG2 identified as both a determinator of CSF Aβ1-42 levels and an attenuator of cognitive decline in Alzheimer's disease. Hum Mol Genet. 2014; 23:1–15. doi:10.1093/hmg/ddu372. [PubMed: 23933734]
5. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson P V, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature. 2012; 488:96–9. doi:10.1038/nature11283. [PubMed: 22801501]
6. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson P V, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med. 2013; 368:107–16. doi:10.1056/NEJMoa1211103. [PubMed: 23150908]
7. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, et al. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013; 368:117–27. doi:10.1056/NEJMoa1211851. [PubMed: 23150934]
8. Bailey-Wilson JE, Wilson AF. Linkage analysis in the next-generation sequencing era. Hum Hered. 2011; 72:228–36. doi:10.1159/000334381. [PubMed: 22189465]
9. Wijsman EM, Pankratz ND, Choi Y, Rothstein JH, Faber KM, Cheng R, et al. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates

CUGBP2 in interaction with APOE. PLoS Genet. 2011; 7:e1001308. doi:10.1371/journal.pgen.1001308. [PubMed: 21379329]

10. Carney RM, Slifer MA, Lin PI, Gaskell PC, Scott WK, Potocky CF, et al. Longitudinal follow-up of late-onset Alzheimer disease families. Am J Med Genet B Neuropsychiatr Genet. 2008; 147B: 571–8. doi:10.1002/ajmg.b.30590. [PubMed: 18361431]
11. Wijsman EM, Daw EW, Yu C-E, Payami H, Steinbart EJ, Nochlin D, et al. Evidence for a novel late-onset Alzheimer disease locus on chromosome 19p13.2. Am J Hum Genet. 2004; 75:398–409. doi:10.1086/423393. [PubMed: 15248153]
12. Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. Genome Res. 2007; 17:1783–6. doi:10.1101/gr.7156307. [PubMed: 17989245]
13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. doi:10.1086/519795. [PubMed: 17701901]
14. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. doi:10.1038/ng1847. [PubMed: 16862161]
15. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. doi:10.1038/ng786. [PubMed: 11731797]
16. Ott J. Linkage probability and its approximate confidence interval under possible heterogeneity. Genet Epidemiol. 1986; 257:251–7.
17. Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics. 1994; 50:118–27. [PubMed: 8086596]
18. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet. 1997; 61:1179–88. doi:10.1086/301592. [PubMed: 9345087]
19. Boehnke M, Ploughman LM. SIMLINK: A Program for Estimating the Power of a Proposed Linkage Study by Computer Simulation. 1997 Version 4.12.
20. Thornton TA, Haines JL. PowerTrim: An automated decision support algorithm for preprocessing family-based genetic data. Am J Hum Genet. 2003; 72:1280–1. doi:10.1086/374823. [PubMed: 12677557]
21. 1000 Genomes Project Consortium. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. doi:10.1038/nature09534. [PubMed: 20981092]
22. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995; 11:241–7. doi:10.1038/ng1195-241. [PubMed: 7581446]
23. Camp NJ, Farnham JM. Correcting for multiple analyses in genomewide linkage studies. Ann Hum Genet. 2001; 65:577–82. doi:doi:10.1017/S0003480001008922. [PubMed: 11851987]
24. Morris JA, Barrett JC. Olorin: combining gene flow with exome sequencing in large family studies of complex disease. Bioinformatics. 2012; 28:3320–1. doi:10.1093/bioinformatics/bts609. [PubMed: 23052039]
25. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet. 2007; 39:17–23. doi:10.1038/ng1934. [PubMed: 17192785]
26. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002; 12:996–1006. doi:10.1101/gr.229102. [PubMed: 12045153]
27. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, et al. The GENCODE pseudogene resource. Genome Biol. 2012; 13:R51. doi:10.1186/gb-2012-13-9-r51. [PubMed: 22951037]
28. Lee JH, Barral S, Cheng R, Chacon I, Santana V, Williamson J, et al. Age-at-onset linkage analysis in Caribbean Hispanics with familial late-onset Alzheimer's disease. Neurogenetics. 2008; 9:51–60. doi:10.1007/s10048-007-0103-3. [PubMed: 17940814]

29. Cronin S, Blauw HM, Veldink JH, van Es M a, Ophoff R a, Bradley DG, et al. Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. Hum Mol Genet. 2008; 17:3392–8. doi:10.1093/hmg/ddn233. [PubMed: 18689356]
30. Chu AY, Guilianini F, Grallert H, Dupuis J, Ballantyne CM, Barratt BJ, et al. Genome-wide association study evaluating lipoprotein-associated phospholipase A2 mass and activity at baseline and after rosuvastatin therapy. Circ Cardiovasc Genet. 2012; 5:676–85. doi:10.1161/CIRCGENETICS.112.963314. [PubMed: 23118302]
31. Hébert SS, Horré K, Nicolaï L, Papadopoulou AS, Mandemakers W, Silahtaroglu AN, et al. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc Natl Acad Sci U S A. 2008; 105:6415–20. doi:10.1073/pnas.0710263105. [PubMed: 18434550]
32. White RE, Giffard RG. MicroRNA-320 induces neurite outgrowth by targeting ARPP-19. Neuroreport. 2012; 23:590–5. doi:10.1097/WNR.0b013e3283540394. [PubMed: 22617447]
33. Saba R, Goodman CD, Huzarewich RLCH, Robertson C, Booth SA. A miRNA signature of prion induced neurodegeneration. PLoS One. 2008; 3:e3652. doi:10.1371/journal.pone.0003652. [PubMed: 18987751]
34. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2012 doi:10.1038/nbt.2450.
35. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011; 25:1915–27. doi:10.1101/gad.17446611. [PubMed: 21890647]
36. Tang B, Di Lena P, Schaffer L, Head SR, Baldi P, Thomas EA. Genome-wide identification of Bcl11b gene targets reveals role in brain-derived neurotrophic factor signaling. PLoS One. 2011; 6:e23691. doi:10.1371/journal.pone.0023691. [PubMed: 21912641]
37. Zuccato C, Cattaneo E. Brain-derived neurotrophic factor in neurodegenerative diseases. Nat Rev Neurol. 2009; 5:311–22. [PubMed: 19498435]
38. Simon R, Brylka H, Schwegler H, Venkataramanappa S, Andratschke J, Wiegreffe C, et al. A dual function of Bcl11b/Ctip2 in hippocampal neurogenesis. EMBO J. 2012; 31:2922–36. doi:10.1038/emboj.2012.142. [PubMed: 22588081]
39. Lazarov O, Marr RA. Of mice and men: neurogenesis, cognition and Alzheimer's disease. Front Aging Neurosci. 2013; 5:43. doi:10.3389/fnagi.2013.00043. [PubMed: 23986699]
40. Ventriglia M, Zanardini R, Bonomini C, Zanetti O, Volpe D, Pasqualetti P, et al. Serum brain-derived neurotrophic factor levels in different neurological diseases. Biomed Res Int. 2013; 2013:901082. doi:10.1155/2013/901082. [PubMed: 24024214]
41. Connor B, Young D, Yan Q, Faull RL, Synek B, Dragunow M. Brain-derived neurotrophic factor is reduced in Alzheimer's disease. Mol Brain Res. 1997; 49:71–81. doi:10.1016/S0169-328X(97)00125-3. [PubMed: 9387865]
42. Weinstein G, Beiser AS, Choi SH, Preis SR, Chen TC, Vorgas D, et al. Serum brain-derived neurotrophic factor and the risk for dementia: the Framingham Heart Study. JAMA Neurol. 2014; 71:55–61. doi:10.1001/jamaneurol.2013.4781. [PubMed: 24276217]
43. Nagahara AH, Merrill DA, Coppola G, Tsukada S, Schroeder BE, Shaked GM, et al. Neuroprotective effects of brain-derived neurotrophic factor in rodent and primate models of Alzheimer's disease. Nat Med. 2009; 15:331–7. doi:10.1038/nm.1912. [PubMed: 19198615]
44. Honea RA, Cruchaga C, Perea RD, Saykin AJ, Burns JM, Weinberger DR, et al. Characterizing the role of brain derived neurotrophic factor genetic variation in Alzheimer's disease neurodegeneration. PLoS One. 2013; 8:e76001. doi:10.1371/journal.pone.0076001. [PubMed: 24086677]
45. Lin Y, Cheng S, Xie Z, Zhang D. Association of rs6265 and rs2030324 Polymorphisms in Brain-Derived Neurotrophic Factor Gene with Alzheimer's Disease: A Meta-Analysis. PLoS One. 2014; 9:e94961. doi:10.1371/journal.pone.0094961. [PubMed: 24733169]
46. Nilsson P, Iwata N, Muramatsu S, Tjernberg LO, Winblad B, Saido TC. Gene therapy in Alzheimer's disease - potential for disease modification. J Cell Mol Med. 2010; 14:741–57. doi:10.1111/j.1582-4934.2010.01038.x. [PubMed: 20158567]

47. Barral S, Reitz C, Vardarajan B, Cheng R, Lee J, Kunkle BW, et al. Linkage analyses in Caribbean Hispanic families identifies novel loci associated with familial late-onset Alzheimer's disease. Alzheimer's Dement. n.d.

48. Chung RH, Schmidt MA, Morris RW, Martin ER. CAPL: a novel association test using case-control and family data and accounting for population stratification. Genet Epidemiol. 2010; 34:747–55. doi:10.1002/gepi.20539. [PubMed: 20878716]

49. Poduslo SE, Yin X, Hargis J, Brumback RA, Mastrianni JA, Schwankhaus J. A familial case of Alzheimer's disease without tau pathology may be linked with chromosome 3 markers. Hum Genet. 1999; 105:32–7. doi:10.1007/s004390051060. [PubMed: 10480352]

50. Liu F, Arias-Vásquez A, Sleegers K, Aulchenko YS, Kayser M, Sanchez-Juan P, et al. A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet. 2007; 81:17–31. doi:10.1086/518720. [PubMed: 17564960]

51. Alzheimer's Disease Sequencing Project (ADSP). 2015. https://www.niagads.org/adsp/content/home

52. Cummings AC, Jiang L, Velez Edwards DR, McCauley JL, Laux R, McFarland LL, et al. Genome-wide association and linkage study in the amish detects a novel candidate late-onset Alzheimer disease gene. Ann Hum Genet. 2012; 76:342–51. doi:10.1111/j.1469-1809.2012.00721.x. [PubMed: 22881374]

53. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R, National Institute on Aging Late-Onset Alzheimer's Disease Family Study G. Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. Arch Neurol. 2008; 65:1518–26. doi:10.1001/archneur.65.11.1518. [PubMed: 19001172]

54. Pericak-Vance MA, Grubber J, Bailey LR, Hedges D. Identification of Novel Genes in Late-Onset Alzheimer's Disease. Exp Gerontol. 2000; 35:1343–52. [PubMed: 11113612]

Systematic review: Pubmed and Google Scholar were used to search for articles related to genetic linkage and genetic association analyses of Alzheimer's disease (AD). Additionally, we searched for literature relating our significant loci to AD and neurodegeneration through a search of each significant chromosomal band (and gene features from the significant multipoint result) and the terms “Alzheimer's” and “Neurodegeneration”. Relevant research relating our significant loci to Alzheimer's or Neurodegeneration is cited.

Interpretation: These findings pinpoint several novel genomic regions linked to increased risk of familial AD, including a region on 14q32.2 containing a gene that regulates brain-derived neurotrophic signaling (BDNF) and the 11q12.2 region previously linked to AD through large genome-wide association analyses of LOAD.

Future directions: Identification of these loci as linked to familial AD provides an exciting opportunity to identify causal variants for LOAD through prioritization of these regions for analyses in forthcoming whole genome sequencing.

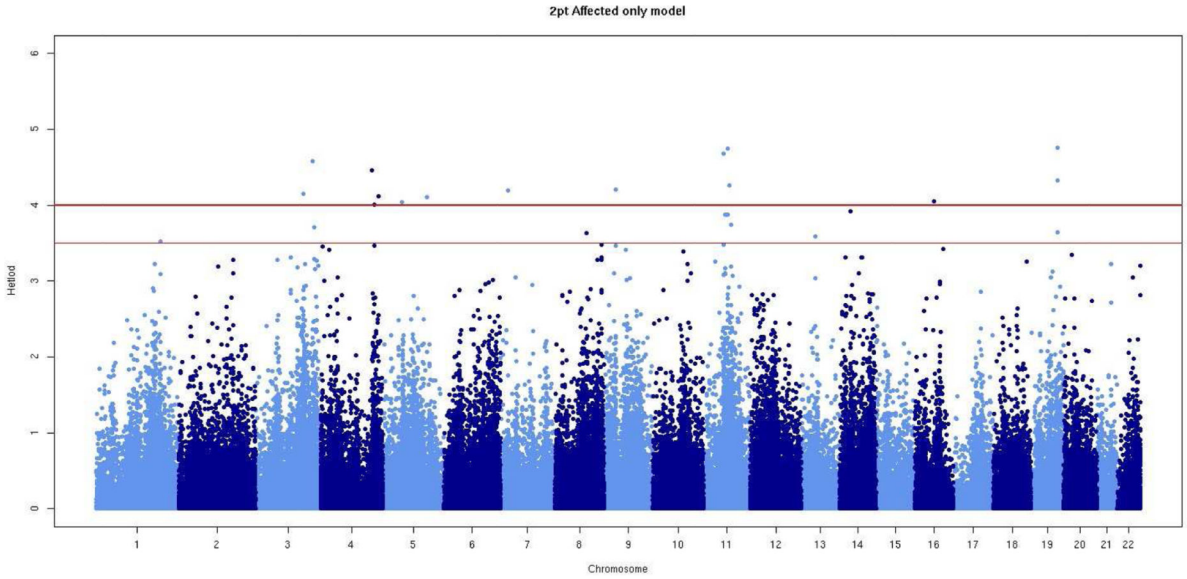

Figure 1.

Manhattan plot of parametric 2-point affecteds-only results (Red lines represent HLOD = 3.5 for significant linkage and 4.0 for highly significant linkage).

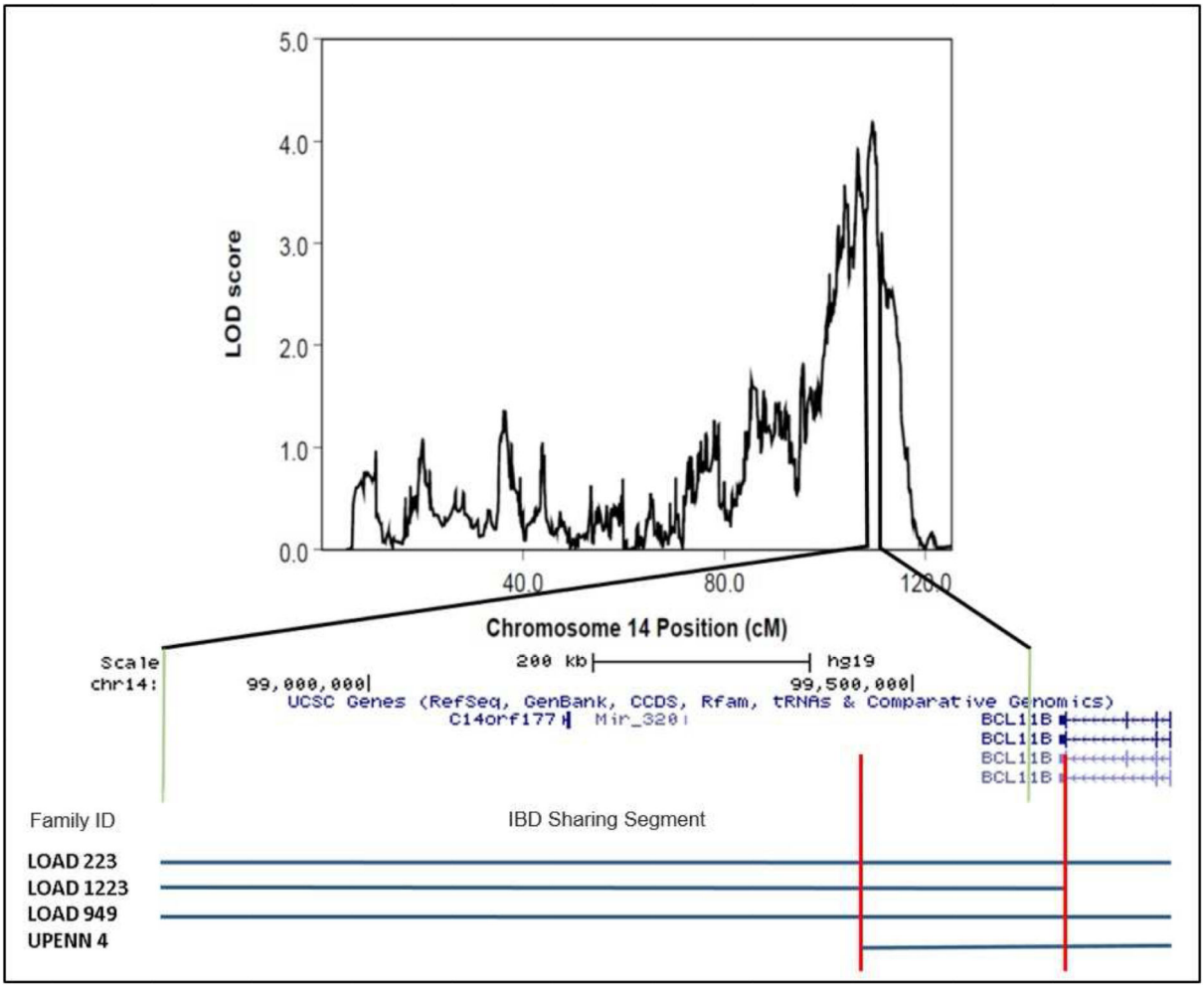

Figure 2.

Shared IBD segments among the four families with LOD* > 0.588 and full IBD sharing for all affected, genotyped family members) in the chromosome 14 linkage region. Red lines represent the minimum shared IBD segment region. Light green lines represent the 1-LOD confidence interval region.

Table 1

Demographic and clinical characteristics of families

Characteristics	
Number of families, n	41
LOAD cases sampled (%)/Unaffected sampled (%)	202 (52.4)/183 (47.5)
≥ 1 autopsy confirmed LOAD case in family (%)	31 (75.6)
Affecteds Sampled Per Family (n families)	
3 Affecteds	7
4 Affecteds	156
5 Affecteds	8
6-9 Affecteds	8
10+ Affecteds	3
Proportion of women, n (%)	246 (64%)
Age at onset of affecteds, years, mean (SD)	75 (SD 9.04)
Age at last examination of unaffecteds, years, mean (SD)	66 (SD 12.23)
APOE Allele Frequency in affecteds, n (%)	
ε2	11 (3%)
ε3	284 (70%)
ε4	109 (27%)

Table 2

Two-point parametric linkage analysis results with HLODs ≥ 3.5.

Chr	BP (Hg19)	Cytogenic Location	RS ID	Ref Allele	Alt Allele	Marker MAF	Affecteds Only Model			Age Penetrance Model			GWAS Locus * (Closest Gene)	Previous Evidence for Region †
							LOD	HLOD	Alpha	LOD	HLOD	Alpha		
1	212028969	1q32.3	rs28451400	C	T	0.07	3.52	3.53	1	2.03	2.03	1	1q32.2 (CR1)	
3a	156034741	3q25.31	rs498033	G	T	0.50	3.70	4.15	0.87	3.99	3.99	1		[52]
3b	186464055	3q27.3	rs6807774	A	G	0.46	4.54	4.58	0.95	2.85	2.85	1		
	187667808	3q27.3	rs2590453	A	G	0.36	3.71	3.71	1	2.52	2.52	1		
4	174854581	4q34.1	rs7670348	C	T	0.18	4.46	4.46	1	1.66	1.66	1		
	179530781	4q34.3	rs1434035	C	T	0.24	4.01	4.01	1	1.61	1.61	1		
	185382061	4q35.1	rs793798	G	A	0.34	4.11	4.11	1	2.52	2.52	1		
5a	41264243	5p13.1	rs391781	A	G	0.42	3.97	3.97	0.93	1.24	1.34	0.82		
5b	145636671	5q32	rs11953090	G	T	0.28	2.43	4.10	0.75	0.00	0.57	0.40		
7	16025340	7p21.2	rs13234986	A	G	0.32	4.19	4.19	1	2.24	2.24	1		[53]
8	101862078	8q22.3	rs4734484	C	T	0.29	3.63	3.63	1	0.93	1.01	0.81		
9	19051208	9p22.1	rs7022613	C	T	0.47	4.21	4.21	1	1.42	1.42	1		[54]
11	60047410	11q12.2	rs1426248	G	A	0.28	4.68	4.68	1	2.40	2.40	1		
	63197930	11q12.3	rs7934347	G	A	0.22	3.85	3.88	0.94	1.44	1.51	0.88		
	73317931	11q13.4	rs737586	A	G	0.17	4.74	4.74	1	2.15	2.15	1	11q12.2 (MS4A6A-MS4A4E)	
	75012535	11q13.4	rs506233	T	C	0.45	3.59	3.88	0.87	1.02	1.35	0.75		
	78879841	11q14.1	rs4359232	T	C	0.42	4.26	4.26	1	1.74	1.74	1		
	83831259	11q14.1	rs7108582	A	C	0.09	3.74	3.74	1	1.70	1.70	1		
13	46047499	13q14.13	rs2985987	G	A	0.14	3.56	3.56	0.93	2.54	2.54	1		
14	36681923	14q13.3	rs1766132	G	A	0.17	3.92	3.92	1	2.05	2.05	1		
16	52112372	16q12.1	rs3743795	T	G	0.32	3.23	4.05	0.84	2.23	2.23	1		

Chr	BP (Hg19)	Cytogenic Location	RS ID	Ref Allele	Alt Allele	Marker MAF	Affecteds Only Model			Age Penetrance Model			GWAS Locus* (Closest Gene)	Previous Evidence for Region†
							LOD	HLOD	Alpha	LOD	HLOD	Alpha		
19	52668414	19q13.41	rs919271	A	G	0.18	4.76	4.76	1	3.23	3.23	1		
	52670905	19q13.41	rs7246914	C	T	0.24	4.33	4.33	1	2.90	2.90	1	CD33 (19q13.41)	
	52699440	19q13.41	rs6509626	C	T	0.10	3.36	4.64	0.87	1.87	1.87	1		

BP = base pairs in GCHr17/Hg19

*GWAS Locus based on Lambert et al. 2013[1]

†Reference for previous evidence of linkage in region based on a significant two-point (LOD ≥ 3.3) or multipoint (LOD ≥ 3.6) score